Forecast Comparisons for Long Memory

Luc Bauwens
SKEMA Business School & CORE, Université catholique de Louvain

Guillaume Chevillon
ESSEC Business School

Sébastien Laurent
Aix-Marseille University (Aix-Marseille School of Economics), CNRS
EHESS, Aix-Marseille Graduate School of Management – IAE

February 28, 2018

Abstract

A large dimensional vector autoregressive (VAR) model can generate long memory in its components under conditions, stated by Chevillon, Hecq and Laurent (2018, CHL), which restrict the VAR parameters. In this context, we compare the forecasting performance of univariate ARFIMA and HAR models, a VAR estimated by ML under the CHL constraints, and a VAR estimated by MCMC. The latter is based on a Gaussian prior density that incorporates the CHL restrictions through the prior mean of the VAR parameters, while the prior variances control the tightness of the restrictions. The forecast comparisons are done on simulated and real data.

Keyword: Long memory, Vector Autoregressive Model, ARFIMA, HAR, MCMC, Forecasting, Volatility.
JEL: C10, C32, C58.

1 Introduction

Long memory is a feature of several types of economic and financial time series, see e.g. ADD REFERENCES. CHL (2015) show that long memory can result from the marginalization of a large dimensional system. More specifically, they provide a parametric framework under which the variables of an n-dimensional VAR(1) can be individually modelled as fractional white noises as n tends to infinity. Long memory may therefore be a feature of univariate or low dimensional models that vanishes when considering larger systems. This source of long memory differs from other sources mentioned in the literature, in particular the aggregation mechanism of Granger (1980).

Our objective is to compare the forecasting performance of several models suitable for series exhibiting long memory. In particular, we consider two univariate models: the autoregressive fractionally integrated moving average (ARFIMA) model, and the heterogeneous autoregressive (HAR) model of Corsi (20??). We also consider a VAR model in large dimension. When its parameters are unrestricted, the VAR model can be estimated by ordinary east squares (OLS) applied to each equation of the system, which is equivalent to Gaussian maximum likelihood. We also use ML estimation of the VAR as a system under different ways of imposing the CHL restrictions, which are parametric restrictions that generate the long memory behaviour of the variables of an n-dimensional VAR.[1] Moroever, we implement Bayesian estimations of the VAR as a system, under a prior density that incorporates the CHL restrictions through the prior mean of the parameters of the autoregressive terms, while the prior variances are chosen to control the tightness of the restrictions. If the prior variances are very large, the prior tightness is null, the VAR model is not parsimonious, and its Bayesian estimation is equivalent to the OLS estimation mentioned above. If the prior variances are set to zero, the restrictions are imposed a priori with probability one, hence the VAR parameterization satisfies fully the CHL restrictions (hence its Bayesian estimation is very close to its ML estimation under constraints)[2] and is very parsimonious, even compared to univariate models. We emphasize that the Bayesian approach is used in this research as a tool to generate a point estimator of the VAR parameters, namely we use the posterior mean as a point estimator.

The motivation for these choices is that the estimation of a univariate ARFIMA model requires a large number of observations to obtain a decent estimate of the long memory parameter. The estimation of an unrestricted VAR model, and a fortiori of restricted versions of it, can be performed with much less observations, even in large dimension. For example, for a VAR(1) in dimension 100, 500 observations is acceptable, while for an ARFIMA on any of the series, this number is too low.

Thus there is a trade-off between the type of model used (from univariate ARFIMA to large VAR, the latter with a increasing number of constraints) and the sample size required for estimation (from large to low). This is the trade-off that we study in this paper. The univariate HAR model is also considered as it captures long memory and does not require as many observations as an ARFIMA model, and is thus more closer to a VAR model with a low lag order, than to an ARFIMA model, in terms of the number of observations required for estimation.

Given that we compare models based on different information sets, and that these models are of reduced form type aimed at forecasting, it makes sense to use comparison criteria based on

[1] In a n-dimensional VAR(1) written as $\mathbf{x}_t = \mathbf{A}_n\mathbf{x}_{t-1} + \epsilon_t$, these restrictions are satisfied if $\mathbf{A}_n$ is a Toeplitz matrix (see, e.g., Gray, 2006) with diagonal elements converging to 1/2 as $n \to \infty$, with off-diagonal elements tending to 0 as $1/n$, and the sum of each row is equal to 1.

[2] Close but not necessarily identical because the posterior mean may not coincide with the MLE, the latter being equal to the posterior mode when the prior is flat.

forecasts. Thus we compare forecasts produced by different models using the root mean squared error (RMSE) criterion as well as the MCS. In addition, we compare the different VAR parameter estimators that we use, in terms of bias and RMSE.

The rest of this paper is organized as follows. Section 2 briefly provides the theoretical framework needed to understand how a VAR can generate long memory, and in particular the parametric restrictions that we use for ML and Bayesian estimation.. Section 3 presents a simulation study designed to quantify the impact of the trade-off between the type of model used and the sample size on estimators and forecasts. In Section **??**, this is illustrated on real data that consist of realized volatilities of forty-nine US stocks. Conclusion are offered in the last section. Technical details are contained in the appendix.

2 Theoretical background

This section summarizes the argument of Chevillon et al. (2018) in the context of the model studied here. We consider a multivariate VAR(p) process, $p \geq 1$, specified as follows. Define $\mathbf{F}_n(L)$, a matrix lag polynomial of dimension n and degree $p-1$, such that the roots of $\det(\mathbf{F}_n(z))$ all have modulus strictly greater than unity. The observable vector $\mathbf{y}_t$ of dimension n satisfies, for $t \geq 1$,

$$(\mathbf{I}_n - \mathbf{A}_n L)\,\mathbf{F}_n(L)\,\mathbf{y}_t = \epsilon_t,$$

where $\epsilon_t \sim \mathsf{IID}(\mathbf{0}, \Sigma_n)$ denotes a process that is identically and independently distributed with zero expectation and variance-covariance matrix Σ_n. Throughout, we assume Σ_n is diagonal with diagonal denoted by $\sigma_n^2 = \left(\sigma_{n,1}^2, ..., \sigma_{n,n}^2\right)$ such that $\sigma_{n,k} > 0$ for $k = 1, ..., n$. Notice that $\mathbf{F}_n(0)$ is not necessarily $\mathbf{I}_n$: this is a parametric representation which corresponds to a non-diagonal error covariance matrix in a standard VAR(p) setting.

In the case consider by Chevillon et al., the matrix $\mathbf{A}_n = \mathbf{T}_n + \eta_n \mathbf{D}_n$, where $\mathbf{T}_n$ is the Toeplitz matrix

$$\mathbf{T}_n = \begin{bmatrix} t_0^{(n)} & t_1^{(n)} & \cdots & t_{n-1}^{(n)} \\ t_{-1}^{(n)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1^{(n)} \\ t_{-(n-1)}^{(n)} & \cdots & t_{-1}^{(n)} & t_0^{(n)} \end{bmatrix}.$$

with entries

$$t_k^{(n)} = \mathrm{Re}\left[\frac{1}{n}\sum_{j=0}^{n-1} g\left(\delta_n, e^{i\frac{2\pi j}{n}}\right) e^{-\frac{2i\pi jk}{n}}\right], \tag{1}$$

where $g(\cdot,\cdot)$ is defined for $\delta \in (0,1)$ and $\omega \geq 0$ as

$$g\left(\delta, e^{i\omega}\right) = 1_{\{0 \leq u < \pi\delta\}} + 1_{\left\{\pi\left(\frac{3}{2}-\delta\right) < u \leq \frac{3\pi}{2}\right\}}, \quad \omega = u \bmod 2\pi, \tag{2}$$

and where $\omega \to g\left(\delta, e^{i\omega}\right)$ is even; the sequence δ_n satisfies $\delta_n = \frac{1}{2} + o\left(n^{-2}\right)$, with $\delta_n < \frac{1}{2}$; η_n is a real scalar sequence that satisfies $\eta_n = o\left(n^{-2}\right)$, $\mathbf{D}_n$ is a real antisymmetric Toeplitz matrix with absolutely summable rows.

Chevillon et al. show that as $n \to \infty$, $(n-1)/4 \in \mathbb{N}$, for all $k \in \mathbb{N}$,

$$\mathbf{F}_n(L)\,\mathbf{y}_t \Rightarrow \Delta^{-1/2}\epsilon_t. \tag{3}$$

In this example, all elements of $\mathbf{x}_t = \mathbf{F}_n(L)\mathbf{y}_t$ within an n-dimensional system present the exact same fractional degree of integration as $n \to \infty$. Since the entries of $\mathbf{A}_n - \frac{1}{2}\mathbf{I}_n$ tend to zero as $n \to \infty$, the cross section dependence between the elements of $\mathbf{x}_t$ vanishes asymptotically. Yet, as $\sum_{k=1}^{n-1} t_k^*(n)$ remains nonzero, the dependence across individual series is sufficient to generate long memory in each of the marginal processes.

3 Simulation and empirical evidence

In this section, we evaluate our key theoretical results via a Monte Carlo simulation. We also show that our theoretical framework is able to replicate some stylized facts observed in the variance of US stock returns.

3.1 Monte Carlo

We provide here simulations that examine the validity of our theoretical asymptotic results when the dimensions of the cross-section and the sample are finite.

An n-dimensional VAR(1), as defined in Equations (**??**)-(**??**), is used to generate data for different choices of T and n. To save space, we only report the results for $n = 201$ series and $T = 4,000$ observations.

As a benchmark, we consider in our first experiment the case of a diagonal matrix, $\mathbf{A}_n = d\mathbf{I}_n$, where the parameter d is set to 0.499. The first panel of Figure 1 shows the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$), i.e., $a_k^{(n)} = 0.499$ for $k = 0$ and 0 otherwise. In this simple setting, the derived univariate processes have short memory and follow a stationary AR(1) model with an autoregressive parameter of 0.499 for each series.

Panel 2 of Figure 1 plots the empirical distribution (over 1,000 replications) of the long memory parameter of series x_{1t} estimated using three popular estimation methods, i.e., the log periodogram regression (GPH) of Geweke and Porter-Hudak (1983), the Local Whittle Likelihood Estimator (LWLE) of Robinson (1995), both with bandwidth $T/2$ and the MLE of an ARFIMA$(1, d, 0)$ (see Sowell, 1992 and Doornik and Ooms, 2004).[3] We deliberately choose a large bandwidth, as implemented by default in Doornik and Ooms (2004) to reduce the variability of the estimators. As expected the estimated long memory parameters are concentrated around 0 suggesting that there is no evidence of long memory in the individual series. This is confirmed by the third panel of Figure 1 which reports the ACF of x_{1t} for the first replication.

In the next two experiments, we consider a symmetric Toeplitz matrix $\mathbf{A}_n = \mathbf{T}_n^*$, under the assumptions of Section **??** (i.e., Equation (**??**) with $\eta_n = 0$), where $\mathbf{T}_n^*$ has symbol g_d. We denote by d the value taken by δ_n : we choose two values of d close to 1/2, i.e., respectively $d = 0.499$ in Figure 2, and $d = 0.45$ in Figure 3. The structure of these figures is similar to that of Figure 1 except that now, since d is close to 1/2, i.e., to the nonstationary region of an $\mathsf{I}(d)$ process, we follow the approach of Abadir, Distaso and Giraitis (2007) and apply the three long memory estimators to $(1-L)^d x_{1t}$ (for the values we report, we have added d ex-post to the estimate). The first panel of these figures emphasizes that the diagonal elements are near d while the off-diagonal elements are small for $d = 0.45$ and very small for $d = 0.499$. Recall from Equation (**??**) that the sum of each row of $\mathbf{T}_n^*$ is unity by construction and therefore although the off-diagonal elements of $\mathbf{A}_n$ can be very small when d is close to 1/2, they are nonzero. Unlike in Figure 1, long memory is detected

[3] All estimations are performed in OxMetrics 7.0 (see Doornik, 2013).

in x_{1t}, with a Monte Carlo mean (over the 1,000 replications) of 0.444, 0.484 and 0.488 respectively for the GPH, LWLE and ARFIMA$(0, d, 0)$ methods for $d = 0.499$ and 0.417, 0.451 and 0.465 for $d = 0.45$. The ACF of x_{1t} in the first replication also suggests the presence of long memory. These figures show that although $\mathbf{A}_n$ is near diagonal, the very small off-diagonal elements play a crucial role in the apparition of long memory.

Next, we evaluate the robustness of the previous result by using the asymmetric Toeplitz matrix given in Equation (**??**), i.e., $\mathbf{A}_n = \mathbf{T}_n^* + \eta_n \mathbf{D}_n$, with $d = 0.499$, $\eta_n = \frac{1}{n^2 \log(n)}$, and where the elements of $\mathbf{D}_n$ in the upper triangle are drawn independently from a standard normal distribution. Figure 4 suggests that results are qualitatively the same as in the case of the symmetric Toeplitz matrix in the sense that long memory is detected in x_{1t} with a parameter estimate close to d.

Theorem **??** states that, under Assumption P, not only x_{1t} but all variables belonging to $\mathbf{x}_t$ should be fractional white noises when $n \to \infty$ and $d \to 1/2$. Our last experiment illustrates this finding for the case of a symmetric Toeplitz matrix with $d = 0.499$, as investigated in Figure 2. Figure 5 plots the empirical distribution of the long memory parameter estimated on all series, i.e., on $x_{1t}, \ldots, x_{201t}$, for the three estimation methods. We only report the results for four replications, each row in the figure corresponding to a replication. Results suggest that the estimated long memory parameters do not vary much across series and are all concentrated in a region close to 1/2, especially for the LWLE and MLE of the ARFIMA$(0, d, 0)$.

3.2 Empirical illustration

The presence of long memory in the volatility is now considered as a stylized fact of the log-returns of financial assets (see Baillie, Bollerslev, and Mikkelsen, 1996, Breidt, Crato, and de Lima, 1998, and Comte and Renault, 1998, among others). As reported in Lieberman and Phillips (2008) "*There is an emerging consensus in empirical finance that realized volatility series typically display long range dependence with a memory parameter d around 0.4 (Andersen et al., 2001; Martens et al., 2004[now 2009]).*"

To illustrate this claim and also to provide a first assessment of the plausibility of our explanation for the origin of long memory, we consider a dataset (provided by TickData) consisting of transaction prices at the 5-minute sampling frequency for 49 large capitalization stocks from the NYSE, AMEX and NASDAQ, covering the period from January 4, 1999 to December 31, 2008 (2,489 trading days).[4] The trading session runs from 9:30 EST until 16:00 EST. Using 5-minute returns, we computed the MedRV estimator of Andersen, Dobrev, and Schaumburg (2012), a non-parametric robust to jumps estimator of the integrated variance.[5]

Figure 6 plots the long memory parameter estimated using an ARFIMA model on $\log(MedRV_{it})$ for $i = 1, \ldots, 49$.[6] The estimated long memory parameters fluctuate around 0.45, with a minimum of about 0.40 and a maximum of about 0.53.

VAR models for the logarithm of realized variances have been used for instance by Anderson

[4]To save space, we do not report company names but only the ticker symbols. There are AAPL, ABT, AXP, BA, BAC, BMY, BP, C, CAT, CL, CSCO, CVX, DELL, DIS, EK, EXC, F, FDX, GE, GM, HD, HNZ, HON, IBM, INTC, JNJ, KO, LLY, MCD, MMM, MOT, MRK, MS, MSFT, ORCL, PEP, PFE, PG, QCOM, SLB, T, TWX, UN, VZ, WFC, WMT, WYE, XOM, XRX.

[5]If $r_{t,i}$ is the ith 5-minutes return of a day t containing M of such returns, the MedRV of day t is computed as $MedRV_t = \frac{\pi}{6-4\sqrt{3}+\pi} \frac{M}{M-2} \sum_{i=3}^{M} med(|r_{t,i}|, |r_{t,i-1}|, |r_{t,i-2}|)^2$, where $med(\cdot)$ denotes the median.

[6]Similar to the previous section, the ARFIMA model is estimated on $(1 - L)^{1/2} \log(MedRV_{it})$ and 1/2 is added ex-post to the estimated value to ensure the estimated d to lie in the covariance stationarity region.

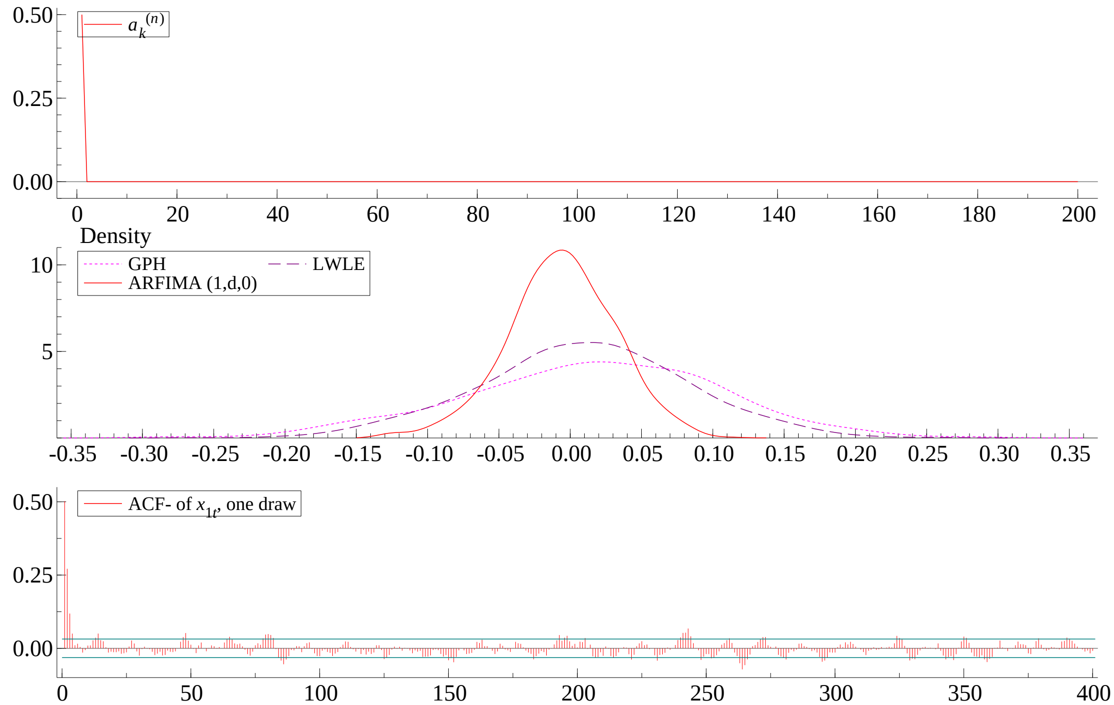

Figure 1: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, with $\mathbf{A}_n = d\mathbf{I}_n$, where $d = 0.499$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The panels report respectively, (a) the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$); (b) the empirical distribution, over 1000 replications, of the estimated long memory parameter of x_{1t} obtained by the GPH, LWLE and ARFIMA$(1, d, 0)$ methods; (c) the empirical ACF of x_{1t} for the first replication.

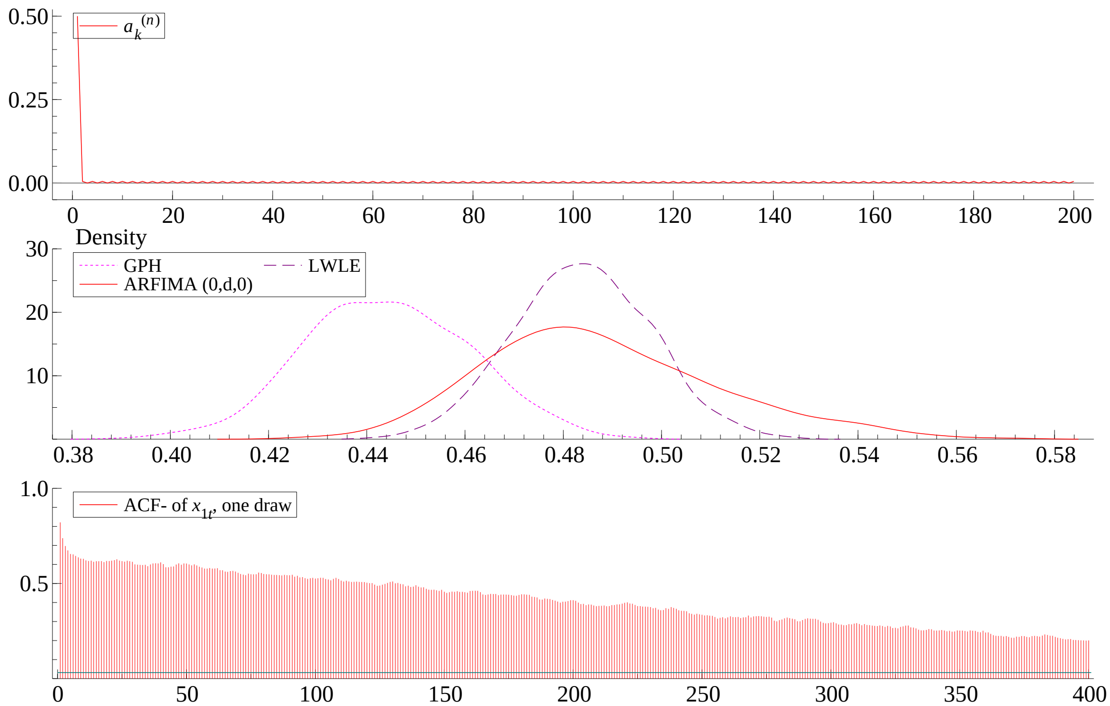

Figure 2: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, with $\mathbf{A}_n = \mathbf{T}_n^*$, where $\mathbf{T}_n^* \equiv \text{Re}\,(\mathbf{T}_n)$, $\mathbf{T}_n$ has symbol defined by (**??**), $d = 0.499$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The panels report respectively, (a) the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$); (b) the empirical distribution, over 1000 replications, of the estimated long memory parameter of x_{1t} obtained by the GPH, LWLE and ARFIMA$(0, d, 0)$ methods; (c) the empirical ACF of x_{1t} for the first replication.

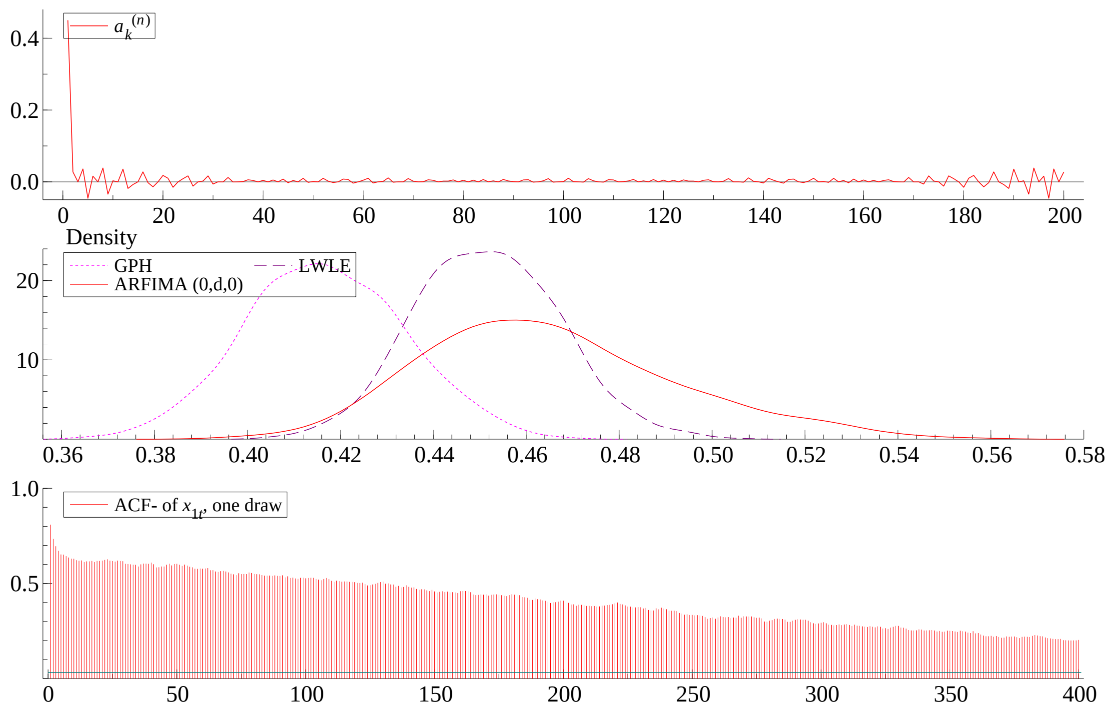

Figure 3: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, with $\mathbf{A}_n = \mathbf{T}_n^*$, where $\mathbf{T}_n^* \equiv \mathrm{Re}\,(\mathbf{T}_n)$, $\mathbf{T}_n$ has symbol defined by (**??**), $d = 0.45$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The panels report respectively, (a) the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$); (b) the empirical distribution, over 1000 replications, of the estimated long memory parameter of x_{1t} obtained by the GPH, LWLE and ARFIMA$(0, d, 0)$ methods; (c) the empirical ACF of x_{1t} for the first replication.

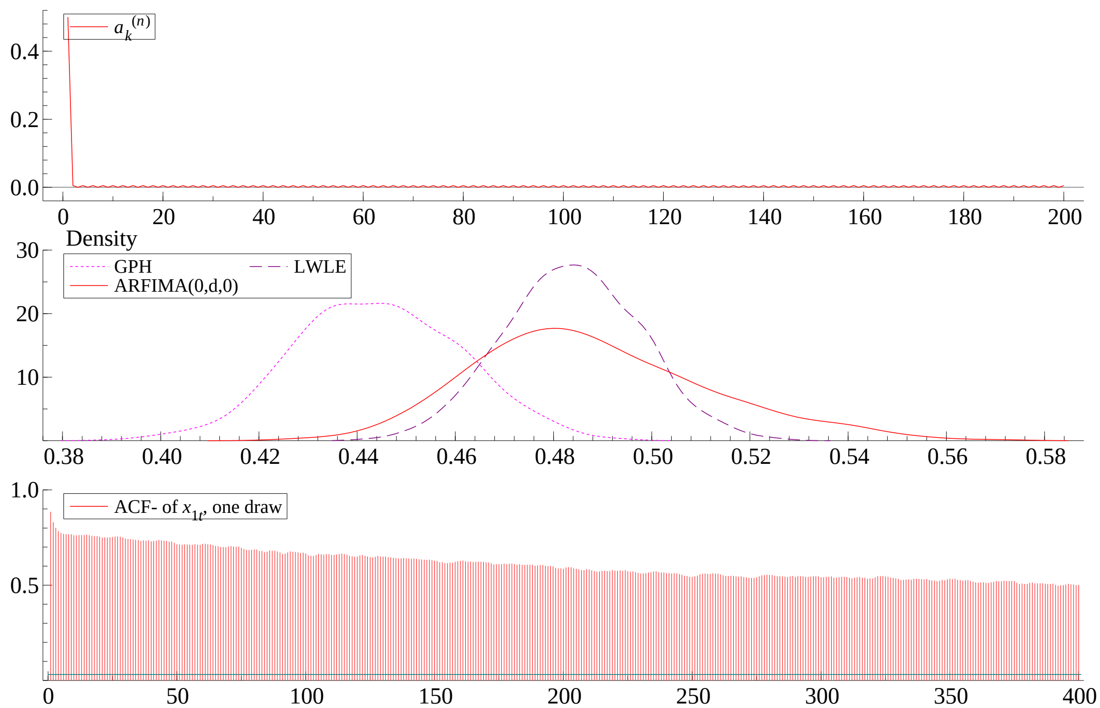

Figure 4: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, with $\mathbf{A}_n = \mathbf{T}_n^* + \eta_n \mathbf{D}_n$, where $\mathbf{T}_n^* \equiv \text{Re}(\mathbf{T}_n)$, $\mathbf{T}_n$ has symbol defined by (**??**), $\eta_n = 1/(n^2 \log(n))$, $\mathbf{D}_n$ is an antisymmetric Toeplitz matrix with above diagonal elements drawn from a standard normal distribution, $d = 0.499$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The panels report respectively, (a) the value of the elements of the first row of $\mathbf{A}_n$, denoted $a_k^{(n)}$ (for $k = 0, \ldots, n-1$); (b) the empirical distribution, over 1000 replications, of the estimated long memory parameter of x_{1t} obtained by the GPH, LWLE and ARFIMA$(0, d, 0)$ methods; (c) the empirical ACF of x_{1t} for the first replication.

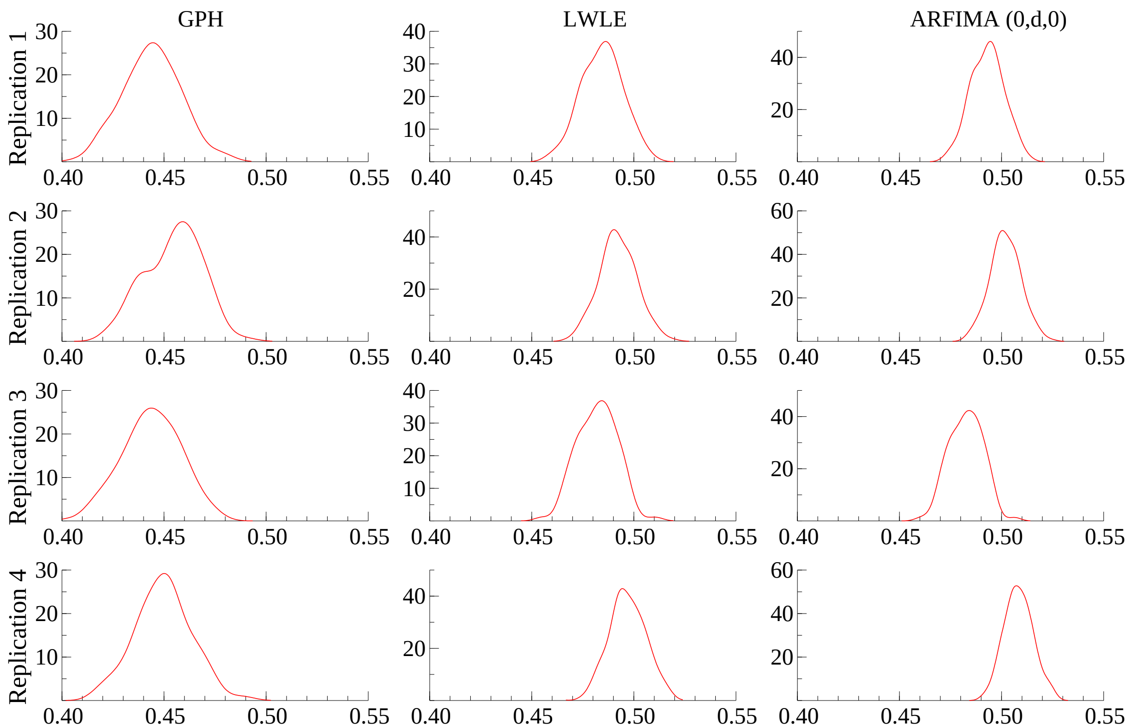

Figure 5: Simulation results for a n-dimensional diagonal VAR(1) $\mathbf{x}_t = \mathbf{A}_n \mathbf{x}_{t-1} + \epsilon_t$, with $\mathbf{A}_n = \mathbf{T}_n^*$, where $\mathbf{T}_n^* \equiv \mathrm{Re}\,(\mathbf{T}_n)$, $\mathbf{T}_n$ has symbol defined by (**??**), $d = 0.499$, $\epsilon_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_n)$, $n = 201$ and $t = 1, \ldots, 4000$. The figure plots the empirical distribution of the long memory parameter estimated on all series, i.e., on $x_{1t}, \ldots, x_{201t}$, using GPH, LWLE and ARFIMA$(0, d, 0)$. Each row corresponds to a separate replication.

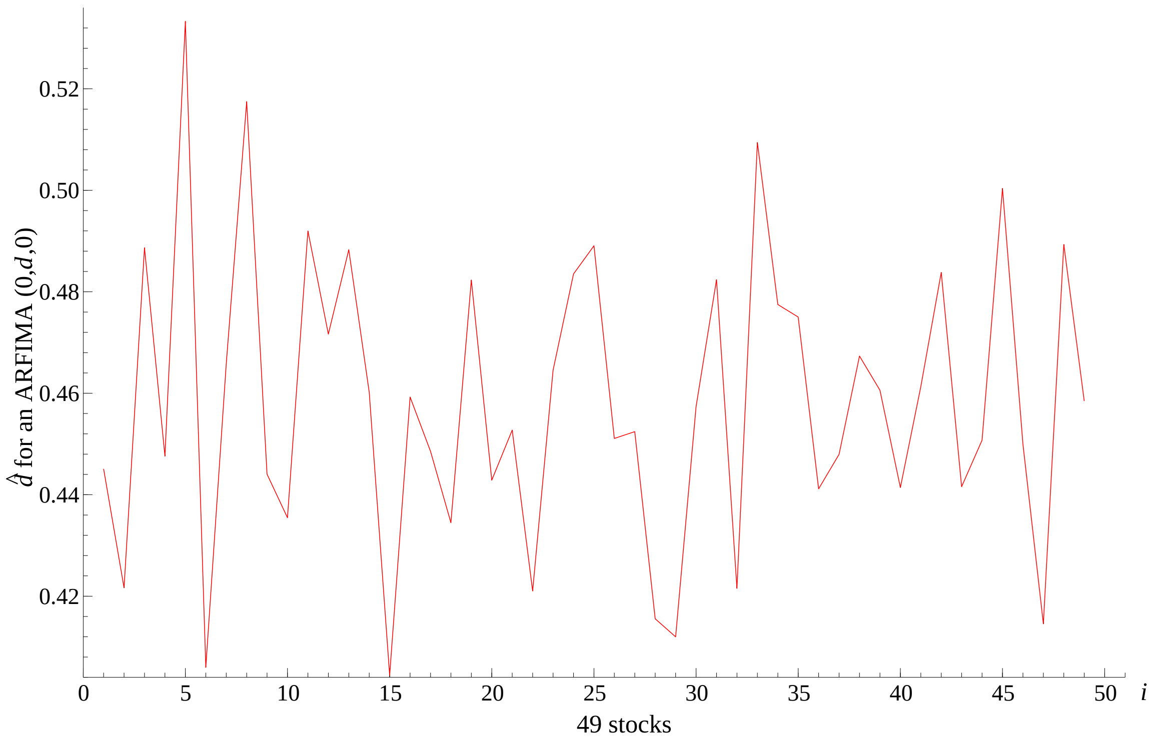

Figure 6: Long memory parameter of an $\text{ARFIMA}(0, d, 0)$ model estimated by maximum likelihood on $\log(MedRV_{it})$ for the 49 stocks $i = 1, \ldots, 49$.

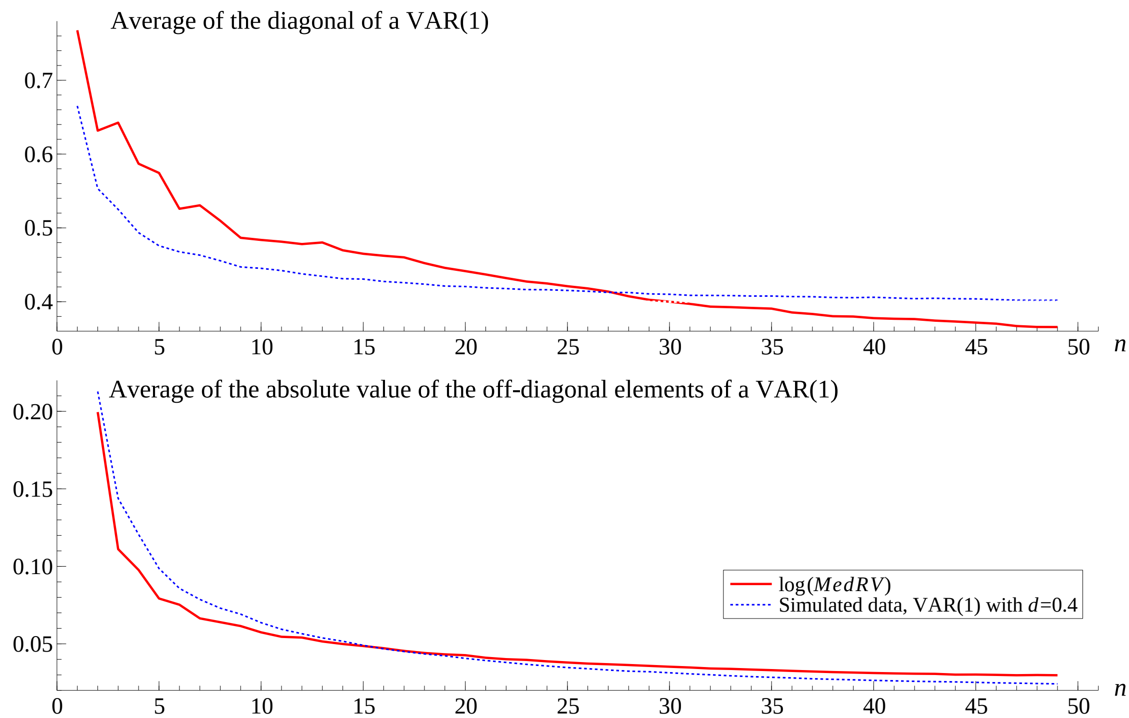

Figure 7: Average of the diagonal elements (upper panel) and average of the absolute value of the off-diagonal elements (lower panels) of a VAR(1) estimated on $\log(MedRV_{it})$ while progressively increasing the dimension of the VAR)

and Vahid (2007). Figure 7 plots some summary statistics on the estimated parameters of a VAR(1) model estimated on $\log(MedRV_{it})$, by progressively increasing the dimension of the VAR (i.e., by adding one variable at a time to the system, following the alphabetical order of the tickers).

The solid lines correspond to the average of the diagonal elements (upper panel) and the average of the absolute value of the off-diagonal elements (lower panel). For instance, the average of the diagonal elements is about 0.63 for the VAR(1) of dimension 2 (i.e., series AAPL and ABT) and the absolute value of the off-diagonal element is about 0.2. Figure 7 suggests that the average of the diagonal elements converges to about 0.4 when the dimension of the system increases while the off-diagonal elements converge to a very small value. This is in agreement with our theoretical model for which the diagonal elements correspond roughly to δ and the off-diagonal elements are small.

Figure 7 (dotted lines) also reports similar quantities but for simulated data following a VAR(1) with a symmetric Toeplitz matrix $\mathbf{A}_n = \mathbf{T}_n^*$, where $\mathbf{T}_n^*$ has symbol g_d given in (**??**), $n = 201$ and $d = 0.4$. While the true dimension of the system is $n = 201$, the VAR is estimated on a smaller system whose dimension progressively increases up to 49 series. A similar pattern is observed both for real and simulated data. Indeed, the average of the diagonal of the VAR(1) estimated on simulated data decreases with the size of the system and converges to 0.4 while the average of the off-diagonal elements converges to a very small value.

References

Abadir, K., W. Distaso, and L. Giraitis (2007). Nonstationarity-extended local Whittle estimation. *Journal of Econometrics 141* (2), 1353–1384.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96* (453), 42–55.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169* (1), 75–93.

Anderson, H. M. and F. Vahid (2007). Forecasting the volatility of Australian stock returns: Do common factors help? *Journal of Business & Economic Statistics 25* (1), 76–90.

Baillie, R., T. Bollerslev, and H. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*, 3–30.

Breidt, F. J., N. Crato, and P. de Lima (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics 83* (1-2), 325–348.

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*, 291–323.

Doornik, J. (2013). *Object-Oriented Matrix Programming Using Ox (7th ed.).* Timberlake Consultants Press.

Doornik, J. A. and M. Ooms (2004). Inference and forecasting for ARFIMA models, with an application to US and UK inflation. *Studies in Nonlinear Dynamics and Econometrics 8*.

Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis 4*, 221–38.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14* (2), 227–238.

Gray, R. M. (2006). *Toeplitz and circulant matrices: A review.* NOW Publishers Inc.

Lieberman, O. and P. C. Phillips (2008). Refined inference on long memory in realized volatility. *Econometric Reviews 27* (1-3), 254–267.

Martens, M., D. van Dijk, and M. de Pooter (2009). Modeling and forecasting S&P 500 volatility: Long memory, structural breaks and nonlinearity. *International Journal of Forecasting 27*, 282–303.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics 23*, 1630–61.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics 53*, 165–188.